# Systems and services sciences: a rationale and a research agenda

David Pym, Richard Taylor, Chris Tofts, Mike Yearworth, Brian Monahan,
Frederic Gittler
Trusted Systems Laboratory
HP Laboratories Bristol
HPL-2006-112
August 14, 2006*

services, services sciences, systems engineering

Services in all of their forms - from consulting to outsourcing - form an increasingly important part of developed economies. Designing and managing efficient and reliable services is not easy, as frequent commercial and public sector failures demonstrate. We must either learn how to specify and manage the complicated systems that services represent or risk the economic consequences of continued failures. In addition to this basic economic imperative there is also the need to provide *secure* systems and services. 'Services Sciences' is being widely pushed as the 'next big thing' for applied research. We propose the *systems* approach to specification, design, implementation, and management as the basis to structure research in this new holistic field.

* Internal Accession Date Only
Approved for External Publication

# Systems and services sciences: a rationale and a research agenda

David Pym, Richard Taylor, Chris Tofts,
Mike Yearworth, Brian Monahan, Frederic Gittler
{*david.pym,richard.taylor,chris.tofts,mike.yearworth,brian.monahan,frederic.gittler@hp.com*}

Hewlett-Packard Laboratories
Filton Road
Bristol BS34 8QZ, UK

August 8, 2006

**Abstract**

Services in all of their forms — from consulting to outsourcing — form an increasingly important part of developed economies. Designing and managing efficient and reliable services is not easy, as frequent commercial and public sector failures demonstrate. We must either learn how to specify and manage the complicated systems that services represent or risk the economic consequences of continued failures. In addition to this basic economic imperative there is also the need to provide *secure* systems and services. 'Services Sciences' is being widely pushed as the 'next big thing' for applied research. We propose the *systems* approach to specification, design, implementation, and management as the basis to structure research in this new holistic field.

## 1 Introduction

> *The greatest enterprise of the mind has always been and always will be the attempted linkage of the sciences and humanities. The ongoing fragmentation of knowledge and the resulting chaos in philosophy are not reflections of the real world but the artefacts of scholarship.*
>
> E.O. Wilson, *Consilience*.

Services in all of their forms — from consulting to outsourcing — form an increasingly important part of developed economies. Designing and managing efficient and reliable services is not easy,

as frequent commercial and public sector failures continue to hamper even some of the most experienced and respected organisations with delivery responsibilities. We must either learn how to specify and manage the complicated systems that services represent or risk the economic consequences of continued failures. In addition to this basic economic imperative, there is also the need to provide *secure* systems and services.

'Services sciences' is discussed with an increasing frequency in both academic and industry communities. The field apparently stretches from anthropology to pure mathematics but lacks a substantive base. There is clearly a need to tie this vast range of 'sciences' together in order to provide confidence that something new will emerge from this new discipline of 'services sciences'.

We want to offer a cohesive approach to the field: the *systems* approach to specification, design, implementation, and management. Placing services in a robust context that enables the analysis and engineering of such systems allows appropriate decisions to be made about research investment and communications.

## 2 Putting the sciences into systems and services

Complicated systems are an integral part of our constructed world. As we come to rely increasingly on these systems for all aspects of our lives we must be able to claim a solid understanding as to how these can be specified, constructed and controlled. With a few exceptions this is not possible. This failure to understand, and hence design and manage complex systems is very evident in the large numbers of high profile information systems failures that we see every year. In an ever more competitive world — whether it be between businesses or nation states, an inability to comprehend, design for, and then deliver effective complex systems is unsustainable. Systems and services sciences are of strategic interest to Europe and the United States. A fundamental problem in understanding, designing and managing real-world complex systems is the need to work fluidly across disciplines. Increasing academic specialisation has tended to work against this — often reflected in industrial research and development as well. Combining disciplines as diverse as psychology, mathematics, and engineering, is difficult. Factors such as language, funding models, publication practices and problem sets all mitigate against the necessary mixture of disciplines coming together to improve our understanding of complex systems, their applications and their limitations.

A conceptual framework for articulating our objectives in systems and services sciences is provided by the scientific method, as depicted in Figure 1. It is perhaps worth restating, for sake of argument, a few basic ideas:

**science**
noun

1. the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment.
2. a systematically organized body of knowledge on any subject.

*Oxford English Dictionary (2005)*

The *scientific method* is the process whereby scientists attempt to create adequate representations of the systems that they are studying — representations that enable reliable and consistent prediction (and hence control) of properties and behaviours. The scientific method has four stages (Figure 1):

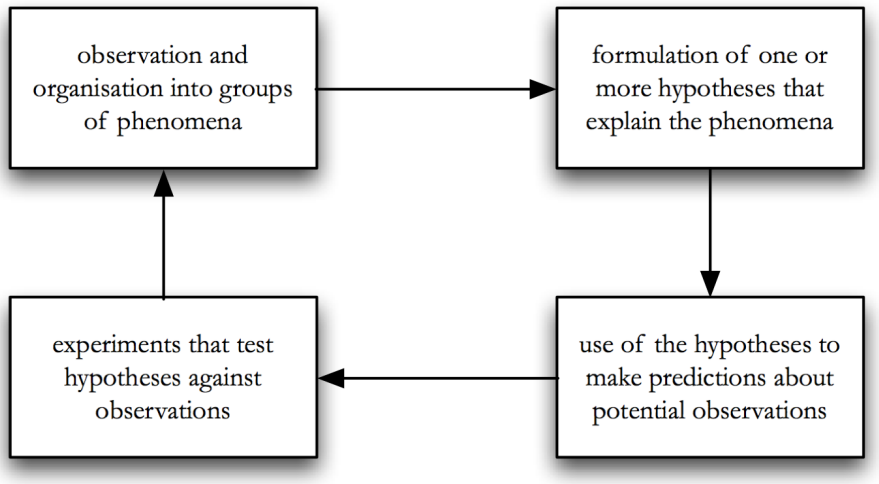


Figure 1: The scientific method in modelling systems and services

1. Observations are made and the observations are ordered by the observed phenomena;
2. One or more hypotheses are formed that explain the observations in way which can be tested and are therefore falsifiable;
3. These hypotheses are used to make predictions about the behaviours of the systems under investigation as they are perturbed;
4. Experiments are carried out that enable the adequacy of those predictions to be tested.

Achieving this closed loop shifts everything, from the scale of a global ITO deal to a dynamically constructed online billing solution for an SME, into the realm of predictable engineering based on sound science. This is the core problem we need to tackle and whether participants are motivated by a desire to develop cost effective, repeatable complicated systems, or to ground academic research or to seek out new problems, disciplines must be fused in a constructive and creative way. The primary purpose is to advance the development and integration of the sciences that underpin the successful analysis, design and control of complex systems characterised by the requirements of services. Grounding these problems in the area of services research — prime examples of large, complicated, and economically significant systems — makes both industrial and academic sense.

We expect that predictions will not be adequate on first use, and so again, expect that the observe-hypothesise-predict-measure cycle will be repeated. It is the ability to predict that makes engineering possible. Any body of research work that does not have as its stated aim the application of the scientific method to the comprehension and construction of services should not be has as much to do with craft as it does with science or engineering.

One apparent consequence of the use of 'sciences' when discussing services research is that this naturally leads to different research camps — economics, computer science, anthropology and operations research to name but a few. This almost immediately raises barriers between groups who need to co-operate, and makes the problem of integrating different bodies of knowledge almost intractable. There is no 'grand unifying theory' for systems and services — or at least not one worth pursuing. It is necessary to think about services sciences slightly differently than most researchers appear to be doing so at the moment if any unification is to occur.

## 3 Services as systems (and not the other way around)

An obvious means of unification is through the reversal of the way that many people are currently looking at the problem. Many conversations assume that from the body of 'services sciences' research, manageable complicated systems will arise. Turn this around. Services are just examples of complicated systems. They can be very complicated indeed, soft requirements such as usability interact with the economic offerings, functional complexity with requirements for systems agility and reusability. The important point is that taking the systems centric view of services enables unification — at least to the extent that it does become possible for social anthropologists to co-exist with mathematicians and control engineers in order to achieve a common goal. Appropriate interfaces between the groups can be placed into the context of the systems under consideration enabling the conversations that need to occur to happen.

As important, treating the service as a system enables appropriate treatment of the system as something to be controlled. This, as we discuss in the following section provides an important philosophical underpinning to the research and its delivery.

### 3.1 The move from Quality Management to Quality Assurance

In the setting of production systems, the change from the build it, apply acceptance test, and then either accept or reject has been profound. The *Quality Assurance* approach of ensure that at *all* stages of production the product is of the required quality greatly reduces the cost of production. This approach is one of the corner stones of lean production. Whilst there are some attempts to describe a 'lean services' approach in the management literature, one of the deep problems in this area is to achieve the necessary approach to quality in services. Without an appropriate approach to quality it is unclear that effective 'lean service' offerings can be developed. In the services setting, the view of quality (SERVQAL for instance) is one of measure, which does permit a quality management approach, not that of *prediction*, which is demanded for a Quality Assurance approach. Consequently, in a complicated service domain the absence of the methods we propose will prove a fundamental barrier to achieving the gains of 'lean services'.

### 3.2 Complicated control spaces - causes and solutions

Examples taken from various publications by the authors:

| *Business as a control system*<br>R. Taylor and C. Tofts | *Self Managed Systems - A Control Theory Perspective*<br>R. Taylor and C. Tofts |
|---|---|
| Businesses can be thought of as control systems;<br><br>The measurement of the businesses behaviour is part of the control loop;<br><br>Understanding the controls the business has is vital;<br><br>Understanding why the measurements were obtained is vital, there **must** be a model;<br><br>Continually checking why deviations from predicitions are occuring, double or triple loop learning. | Control systems are becoming distributed;<br><br>But they still must have the properties of **good** control systems;<br><br>Consequently we must be able to predict their behaviour over their complete domain of operation;<br><br>Simulation will never predict over complete domain with any reliability;<br><br>Either need to do the proofs of the control system correctness or have systems which are good by construction. |
| *Constructing Stable Control Systems Using Cellular Automata*<br>R. Taylor | *Reductionism isn't Functional*<br>M. Hatcher and C. Tofts |
| Fully expressive cellular automata are Turing complete;<br><br>Therefore their behaviour cannot be predicted in general;<br><br>Distributed control systems can be thought of a systems of general cellular automata;<br><br>To get stable control from such systems we must restrict to automata classes on which we can predict that stable behaviour is inevitable. | Synchronous process algebra can capture the computational behaviour of arbitrary compositions of arbitrary contingent systems;<br><br>Its extensions to time, probabilistic and priority behaviours may permit the same over these domains;<br><br>We can prove properties of these systems, if they are small enough;<br><br>We can certainly formally represent them within a small language;<br><br>Currently this space is not calculationally effective at scale. |

# 4 Crisis? What crisis? Research in a post-artefact world

## 4.1 Research as it used to be done

As an example consider how physics and chemistry research required the mathematics necessary to develop hardware (sensors, integrated instruments, etc.) for experimental work to be conducted. A demonstration consisted of plugging something in and watching something happen — it was clear (often) where the individual contribution was being made. Technology transfer occurred through a combination of artifact transfer, documentation, and often a considerable amount of co-development (basing members of the research team with the R&D team). Just as much of HP's R&D was characterised as 'next-bench' design, much of HPL's successful research was 'next-bench' research.

Then we began to write software, and the development of software as instantiations of smart IP became slowly accepted as an acceptable deliverable. This was not an entirely painless transition of mindset. Again, the transfer process was similar to that of traditional research products.

While this remained a next-bench activity (operating systems, compilers, design tools, network monitoring systems are all good examples), it provided a (generally) well focussed path from idea through to product. Then, as software became a more important part of HP's offerings, the 'next bench' paradigm began to break down. One well-known example of an attempt to break out of 'next bench' was e-speak.

## 4.2 A third age for research and development

Finally, to the third age — that of *research for services*. In many ways this resembles the transition from hardware only artifacts to hardware and/or software research deliverables. There is the same reluctance to change modes of research and delivery interactions by potential research staff, issues to do with the encouragement of appropriate behaviours through reward and recognition and the need for a genuine debate over exactly what does research for services mean to our organisation?

The most obvious issue is that of next bench design and delivery. Laboratory experiments to investigate and demonstrate large systems behaviour are difficult to source. Conducting the same experiments in the on customers is fraught with dangers. These systems must be (otherwise they should not be paid for) integral to their core business, leaving them extremely vulnerable to the economic and social consequences of failure. There have been standing jokes about certain software houses treating their customers as beta or even alpha test sites for their products. This can never be acceptable for services, although some equivalent to informed consent for co-development might work for both sides. One way or another, researchers need to take a different approach to cooperative research which brings them very much closer to the customer for at least some of the work.

The second major difference is that now research deliverables are not just a block of software, a hardware design or a research paper. For most projects in this space we will expect deliverables to be some mixture of

- theoretical and experimental analyses of aspects of systems behaviour and construction (conventional),
- embodiments of theoretical results in the form of tools for management, analysis and planning (conventional),
- embodiments of theoretical results in the form of processes (not something that many traditional research laboratories have practiced), and
- training, for both in-company audiences — the people who will design and deliver services — customers (to enable them to make best use of our offerings), and a wider audience of technologists, including students who may come to work for either ourselves or our customers (some tradition but not well-articulated or part of a wider strategy).

If this is to occur, we need to define a broader group of responsibilities for research staff as well as extending relationships with Universities, including groups with whom we have traditionally not seen as affecting our future businesses.

For example, HP currently has a model-based analysis and design methodology called 'Rapid Scenario Planning' (RaSP). Briefly, this enables disparate groups of stakeholders in a system project to improve their understanding of the couplings between their concerns, resolve (in as far as possible) primary inconsistencies and then track these as the project progresses. Developed initially for e-service projects, this has become part of HP's *Open Analytics* offerings. Traditionally this type of process-technology development has not been a significant part of HP Laboratories' work, and it has not been clear how we could make use of academic resources to develop it further. Several potential answers have emerged:

1. Work with a 'scenario planning' research group from a major business school both to improve our understanding of the sociology of scenario planning and to enable them to understand better the use of more formal mathematical methods to manage scenario planning (the quid pro quo);
2. Work with a project management research group, in this case based across business and computer science schools, to train masters level students in the basics of RaSP, enabling them to use RaSP with the subjects of their dissertations (generally medium enterprise-sized customers) as experiments to enable a better understanding of both the skills necessary to deliver the process and the broader benefits to that understanding;
3. Work with a business modelling research group to run experiments on the impact of representation techniques for different stakeholders, using MBA-level students (all of whom will have held middle-ranking appointments in business or government) as subjects.

None of these suggestions represent, on their own, radically different approaches to driving forward research in this particular area, but taken as a whole they do represent a change in the way, for example, that we are approaching potential academic collaborators.

## 5 The scope of the research

Amongst the many potential research, development and delivery problems that could become part of this group of projects, we propose focussing on the list below. The diversity of this proposal reflects the nature of the problem: we must integrate, social, economic, engineering, and mathematical sciences in pursuit of a coherent collection of systems and services sciences.

- Developing the underlying mathematics of systems and services sciences — all science ultimately depends on solid foundations, and we must encourage the development of appropriate mathematics — presentation, analysis and explanation that will enable the gulf between the 'general' and the 'specific' to be bridged. Engineering (in all of its forms from mechanical to electrical) provides a model for the repeatable and reliable application of mathematics and science, and we propose that this should be repeated for systems and services
- Dealing with socio-technical integration of services: the interfaces between 'hard' and 'soft' technologies, as well as abilities to reason between the 'why' and the 'what' are poorly understood. Specifically, we propose projects that examine:
  - Composing people, processes and technologies — how can we treat systems that have social requirements, driving organisations, driving processes which in their turn drive information systems provision?
  - Relationship management; dealing with systems of systems - the grounding of management science approaches to 'systems' in concrete and reusable forms that have a rational mathematical basis.
- Development, validation and maintenance of direct relationships between the economic, financial, human and technical properties of the systems under discussion.
- Developing rigorous, composable, reusable and comprehensible systems modelling tools — there are many approaches to systems analysis which are difficult to compare and contrast. A shared understanding, with appropriate tools and representation standards, will enable purchasing organisations to make rational and repeatable choices between providers.
- Educating and training of personnel across industry — all industries need to be able to train to recognised standards, both for their own benefit and that of their customers. In the area of systems and services sciences, the establishment of appropriate training and the recognition of the validity of qualifications is neither agreed upon nor well-articulated. This activity will establish both.

- Developing the underlying mathematics of systems and services science; tackling intermediate scale modelling problems (cf. climate modelling for weather forecasting — extremely hard). Appendix 1 contains an outline example of a possible research project.
- Integrating modelling methods for systems of discrete components with methods that model large scale dynamics.
- Developing the scientific infrastructure to support systems and services sciences as an academic discipline - initiatives including the cross industry Centre for Systems and Services Sciences (CS3) point the way to the effective development of both social and technical infrastructure. Specifically it is essential that there is support for the following:
  - Organisation of regular hybrid industrial-government-academic research meetings and colloquia;
  - Publication of advances in the area of Systems and Services Sciences through a refereed journal;
  - Identification and encouragement of pre- and near-competitive research
  - Establishment of a 'dating agency' for academic, government and industrial partnerships;
  - Establishment of systems and services sciences as a recognised research discipline;
  - Establishment and championing standards in systems analysis, specification, development and management in process and training;
  - A body that can act as a validator of expertise of competence in systems and services sciences;
  - A body that can act as a repository for sample and standard problem sets and research results.

## 6 Conclusions

'Services Sciences' has become a confused and ill defined term. With poor definitions of what it means, defining and executing a research agenda in this space is difficult. Concentrating on 'Systems and Services Sciences' — i.e., the understanding of how systems operate and how that understanding can be used as an integrative framework for the multiple 'sciences' that underpin services is practical.

We intend on moving forward in this research area, extending beyond traditional means of interacting with both academia, partners, competitors and the public sector in the area of services standards and research.

# A An example of a collaborative research programme: hybrid modelling technologies for systems and services sciences

***Authors: David Pym, Richard Taylor, Chris Tofts, Mike Yearworth***

## A.1 This Document

This document is an example of a research proposal in the area of systems and services sciences. Its focus is on the mathematical and conceptual foundations of modelling technologies.

For simplicity and brevity, we omit the usual scholarly referencing and citation.

## A.2 Introduction

The mathematical modelling of physical and economic systems has a long and distinguished history of achievement, with fundamental developments in mathematics developing hand-in-hand with increased understanding of the world being modelled.

On the one hand, the use of continuous mathematics, primarily differential equations and related topics, has allowed not only the development of theoretical physics, but also has allowed engineers to build our physical environment. On the other, the behaviour of complex communication systems can be understood using information-theoretic, probabilistic, and statistical methods.

In recent decades, the world has become dependent upon large, complex computer systems. Yet the design and delivery of these systems is widely understood to be highly problematic: implementation delays, inadequate performance, unreliability.

*Recent work by Taylor, Tofts, and Yearworth at HP Labs, Bristol has demonstrated how the design of contracts (e.g., Service Level Agreements or SLAs) between the suppliers and users of computer systems and the design of the computer systems being supplied should be co-developed in order to achieve enhanced effectiveness and reliability.*

From another perspective, techniques from mathematical logic and theoretical computer science have been used as bases for understanding the compositional structure of complex assemblies of discrete components, with applications to topics such as system specfication and verification. Indeed, *Recent work by Pym and Tofts, O'Hearn, Reynolds, Yang, and others, on logics, programming languages, and process calculi* has demonstrated that bunched logic and its relatives, such as separation logic , and SCRP-MBI (see below) can provide elegant solutions to previously rather poorly handled problems.

Taking a systems-level perspective, it is evident that we require a mathematical framework that allows these diverse levels and types of models to be related to each other.

This proposal is to develop a mathematical framework, together with supporting computer-based tools, for modelling the architecture and dynamics of complex assemblies of discrete components so as to support

- Precise specifications of the structure and behaviour of critical components of systems.
- Discrete, compositional models of the structure and behaviour of components systems and of whole systems. For example, techniques from queuing theory and probability theory have been in tools, such as DEMOS2k, used at HP Labs to model the performance of large IT systems.
- Continuous models of the behaviour of systems, such as the performance of a network or a collection of servers, or models of interacting species as represented by systems of differential equations.
- A precise understanding of the interaction between discrete and continuous representations.
- Support for the co-design of contracts and systems relative to chosen models.

An important aspect of all of these requirements will be the necessity of maintaining the ability to set up models at levels of abstraction appropriate to the understanding and information required. For example, in order model the reliability of a multiprocessor architecture, there is no need to model the internal structure of each type of microprocessor if one has the reliability data for each type of processor.

## A.3 Business Relevance and Context

### A.3.1 Systems projects

The failure to meet expectations (be they cost, time to delivery, function, performance or reliability) of large systems projects is depressingly familiar. Failures are not confined to public sector systems (even if these are more immediately obvious), but the increasing (almost pervasive) use of public-private sector partnerships is exposing the complete 'systems' industry lack of mechanisms for understanding and controlling these projects. Failure can be catastrophic for all parties involved, with the customer losing business (or in the case of government, control and credibility), and the supplier very large sums of money. In the UK alone, as much as 40 of all new information systems investment can be considered wasted, with a further 30 consumed by cost and time over-runs.

There are many reasons why projects fail to meet their users expectations. Commonly identified culprits include:

- apparently overwhelming systems complexity — people, processes, information systems and financial constraints;
- failures in the establishment and understanding of system specifications;
- failures to identify commonality amongst projects - a tendency to view each and every project as something unique, restricting re-use;
- the closed nature of many information systems projects whereby catastrophic failure is not frequently used to inform other projects (unlike other engineering disciplines such as civil engineering) due to commercial constraints;
- poor education and common standards for 'whole-life' systems analysis.

Industry will argue that it has made great progress in the establishment of common standards for systems specification and project management (SCOR, ITSM, and Prince, for example). These techniques fail to address a primary reason for systems failure – a lack of comprehension of the system that must be constructed and operated. This lack of adequate comprehension acts as a significant brake on our abilities to reliably and repeatedly design and manage large complex systems.

### A.3.2 Industrial experience of model based analysis and project management

Hewlett-Packard, through an extensive research and development programme, has been experimenting with the use of modelling technologies, with leading examples being DEMOS2k and related tools, as an integral part of commercial project development and management. Our conclusions about the effectiveness of model-based approaches can be summarized as follows:
- the act of creating a model forces an organization to consider and review the structure of the business, investment or product that they are proposing to create; such a model, even if it is never deployed in anger or formally analysed will often play an important role in the initial feasibility study;
- the model acts as important part of the documentation of a system; the evolution of such a model, if documented, is an invaluable aid in the audit of a project;
- the model can act as a valuable communications aid, allowing discussions to be grounded in a common representation;
- models allow for rapid exploration of the decision space that an organisation is operating in, enabling multiple scenarios to be played out at low risk;
- models may be used to qualify and then check real systems; as the system runs, the behaviour of the model is compared with observations of the real system and discrepancies are investigated;
- models can demonstrate the sensitivity of a system to environmental changes, enabling users to design out (as much as is possible) potentially disruptive non linearities in the system behaviour;
- models can be used to check the correctness of particular approaches to problem solving;
- models permit the early capture of error, as they permit nonexistent systems to be studied, with the well known benefit of capture time against value saved.

An essential aspect of all of these requirements is the is the need to maintain the ability to establish models at appropriate levels of abstraction that match both the level of required detail and the quality of analytic results. For example, a 'system' in the business sense will consist of objectives that are served through processes, people, and infrastructure. In order to be able to design and manage the system it is necessary to understand the impact of structure and behaviour on the static and dynamic properties of the collective, but the level of detail and precision will vary dependent upon the questions being asked. When reasoning about the reliability of a particular process and the business impact of shifting reliabilities it is not necessary to model the hardware infrastructure at component level.

The primary vehicle for exploring the impact of model based design and management has been a process known as 'Rapid Scenario Planning' (RaSP) based upon tools that enable rapid performance and availability analysis of system processes and resources. The primary intention of RaSP is to reduce stakeholder disagreement, identify areas of maximum sensitivity to initial assumptions and plan systems specification, design and operation.

While still comparatively crude compared to the work that is being proposed in this document, RaSP and it's associated analyses have been successfully applied to more than £8B worth of complex systems projects in the UK, the wider EU and the US.

### A.3.3 Pre-competitive research

It is important to emphasise that markets only exist where there is competition. In order for integrated complex systems modelling and analysis to become more widely used within business several things need to happen

- agreed standards for modelling techniques and audit;
- open access to common forms of representation that can be exploited across different industries and by both competing and cooperating bodies.

If it is possible for industry and academia to come to some agreements over these, then competition can be based upon the relative efficiencies of their application - essentially coming down to the abilities of analysts, the qualities of their model libraries (which make up competitive research fodder), their organization and their abilities to apply the findings to business systems.

### A.3.4 Commercial potential

The primary commercial potential of this work does not lie in the creation of new businesses that will grow from research development and/or breakthroughs. While new businesses will undoubtedly

be enabled, their value will be dwarfed by the potential savings that this work can create, measured by improved success rates in the delivery and management of complex systems. In the UK alone, a 10 improvement in efficiency will generate at least £1.2Bn per annum.

## A.4 Scientific Background

We begin by describing our background theoretical work on modelling resources and processes.

General purpose simulation languages allow us to represent and study all of the above problems. They can be particularly beneficial for many reasons including

- programmatic: consequently many users,
- dynamic — give a feel for the system behaviour,
- interactive — can change system properties 'on the fly' to see the effect,
- immediate — can often be constructed very quickly from library components,
- concrete — do not require abstraction of the system under study.

DEMOS is a good example of a simulation environment in this class. However, simulation has many faults, including

- computationally inefficient — usually compared with analytic solutions,
- difficult to scan over variables,
- very costly to seek optima or flat regions in the performance space,
- exposed to experimental risk,
- need to be careful of 'random' number generation,
- no need to clearly identify the problem — too concrete.

Despite all of these defincencies, simulation-style languages are a good approach to problem-capture. So, one approach is to develop the mathematics that should underpin them. Recent work on the logical and process-algebraic modelling of resources and the interaction between resources provides a mathematical basis for this endeavour.

The notion of *resource* is a basic one in many fields, including economics, engineering, and the humanities, but it is perhaps most clearly illuminated in the computing sciences. The location, ownership, access to, and consumption of resources are central concerns in the design of systems,

such as networks, within which processors must access devices such as file servers, disks, and printers, and in the design of programs, which access memory and manipulate data structures, such as pointers.

Within mathematical models of computational systems, however, the rôle of resource is quite central. This observation is illustrated quite directly in modelling systems such as DEMOS, in which the central notions are *entities*, which execute trajectories within the model, and *resources*, which are required to enable, and are manipulated by, entities' actions. The semantics of DEMOS, however, relies on a purely process-theoretic representation of resource. We would argue that this situation is conceptually unsatisfactory. Moreover, pragmatically, the computational cost of modelling interactive systems is, typically, dominated by the handling of the resource components.

A mathematical account of a useful notion of resource can be given using logic. Our starting position is that the following properties are reasonable requirements for a simple model of resource:

- A set $\mathbf{R}$ of resource elements;
- A (partial) combination, $\circ : \mathbf{R} \times \mathbf{R} \rightharpoonup \mathbf{R}$ of resource elements;
- A comparison, $\sqsubseteq$, of resource elements; and
- A zero resource element, $e$.

In the usual spirit and methodology of mathematical modelling, these conceptually evidently well-motivated properties correspond well to a wide a range of natural examples. Mathematically, we obtain this structure as pre-ordered partial commutative monoid, $\mathcal{R} = (\mathbf{R}\,,\,\circ\,,\,e\,,\,\sqsubseteq)$, subject to the condition that if $r \sqsubseteq s$ and $r' \sqsubseteq s'$, then $r \circ r' \sqsubseteq s \circ s'$, and, recalling the preordering of a Kripke structure, call it a *Kripke resource monoid*, or KRM, with worlds being resources. The ordering $\sqsubseteq$ gives rise to an equality, $= \;\; := \;\; \sqsupseteq \cup \sqsubseteq$.

A simple example is provided by the natural numbers, here including 0,

$$\mathcal{N} = (\mathbb{N}\,,\,+\,,\,0\,,\,\leq),$$

in which combination is given by addition, with unit 0, and comparison is given by less than or equals. This is an example of resource as *cost*. Many more examples may be found in .

Kripke resource monoids provide the basis for the semantics of **BI**, the logic of bunched implications. The judgement $r \models \phi$, for $r \in \mathbf{R}$, is read as 'resource element $r$ is sufficient to support proposition $\phi$'. The ordering structure admits the usual Kripke semantics for the usual, additive, connectives $(\top,\ \wedge,\ \bot,\ \vee,\ \rightarrow)$ of intuitionistic logic and, in the discrete case, classical logic.[1] The monoidal structure admits a semantics for a multiplicative conjunction, $*$, given by

$$\begin{array}{rcl} r \models \phi_1 * \phi_2 & \text{iff} & \exists\ s_1 \text{ and } s_2 \text{ s.t. } s_1 \circ s_2 \sqsubseteq r, \text{ and} \\ & & s_1 \models \phi_1 \text{ and } s_2 \models \phi_2. \end{array}$$

[1] Our use of the terms 'additive' and 'multiplicative' derives from their use in linear logic and bunched logic.

The semantics of the multiplicative conjunction, $*$, is interpreted as follows: the resource $r$ is sufficient to support $\phi_1 * \phi_2$ just in case it can be divided into resources $s_1$ and $s_2$ such that $s_1$ is sufficient to support $\phi_1$ and $s_2$ is sufficient to support $\phi_2$. The assertions $\phi_1$ and $\phi_2$ — think of them as expressing properties of programs — *do not share* resources. In contrast, in the semantics of the additive conjunction, $r \models \phi_1 \wedge \phi_2$ iff $r \models \phi_1$ and $r \models \phi_2$, the assertions $\phi_1$ and $\phi_2$ may *share* the resource $m$.

Along with the multiplicative conjunction comes a multiplicative implication, $-\!\!*$, given by

$$r \models \phi -\!\!* \psi \quad \text{iff} \quad \begin{array}{l} \text{for all } s \text{ such that } s \models \phi, \\ \quad r \circ s \models \psi. \end{array}$$

The semantics of the multiplicative implication, $-\!\!*$, may be interpreted as follows: the resource $r$ is sufficient to support $\phi -\!\!* \psi$ just in case for any resource $s$ which is sufficient to support $\phi$ the combination $r \circ s$ is sufficient to support $\psi$.

We can think of the proposition $\phi -\!\!* \psi$ as (the type of) a function and the proposition $\phi$ as (the type of) its argument. The resources then describe the cost of applying the function to its argument in order to obtain the result. The function and its argument *do not share* resources.

The composition and ordering structure lifts to sets of resource elements, to give what we might call a basic separation model. Let $\wp(\mathbf{R})$ denote the powerset of $\mathbf{R}$ and let $R, S \in \wp(\mathbf{R})$. Then define, for example,

$$R \circ S = \begin{cases} \{r \circ s \mid r \in R,\, s \in S\} & \text{if each } r \circ s \downarrow \\ \uparrow & \text{otherwise,} \end{cases}$$

with unit $\{e\}$ and, for example,[2]

$$R \sqsubseteq S \quad \text{iff} \quad \forall\, r \in R\ \exists\, s \in S \text{ s.t. } r \sqsubseteq s.$$

Such sets of resources are a convenient level of abstraction for our present purposes, for which we shall require no further special properties. We might also require that $R \circ S$ be defined only if $R$ and $S$ are disjoint. We write $R_1, R_2$ for the union of $R_1$ and $R_2$, and emphasize that composition is quite different from union. Our notational choices should be clear *in situ*. Other constructions, based on Kripke resource monoids, might also provide a basis for a calculus and logic. The space of choices is, however, quite large, so that a discussion of it is beyond our present scope.

More generally, we might take a more complex structure of resources. For example, we might take $\mathbf{R} = \mathbf{R}_1 \times \ldots \times \mathbf{R}_m$, with a composition $\circ_i$ and ordering $\sqsubseteq_i$ on each $\mathbf{R}_i$.

We emphasize that **BI** and linear logic are very different. Logically, they are incomparable extensions of a basic system, sometimes called Lambek logic, of a (commutative) tensor product, with a unit, and an implication; their treatments of the additives and the structural laws of weakening and contraction are radically different. Moreover, linear logic's resource reading amounts to counting

[2]Note that the ordering on $\wp(\mathbf{R})$ given here is just one of many possible choices.

occurrences of propositions, whereas **BI**'s resource semantics incorporates a basic model of the notion of resource.

Returning to our earlier 'basic separation model', we note that taking the ordering to be given by equality gives a model of Boolean **BI**, that is, with classical additives. (A model of **BI** with intuitionistic additives is obtained by taking $\sqsubseteq$ to be inclusion.) In this model, if $\phi * \psi$ holds for a given collection of resources, then $\phi$ and $\psi$ hold for disjoint sub-collections.

A starting point for the development of a more satisfactory account of the relationship between resource and process has been provided by the logic **BI** and its various formulations , as discussed in § A.4, together with Milner's synchronous calculus of communication systems, SCCS, which provides a beautiful operational (proof-theoretic) account of the evolution of concurrent processes. This work has led to a large thriving school of work in process algebra and the theory of concurrency.

We have developed a calculus, called **SCRP**, in the spirit of Milner's synchronous calculus for communicating systems, SCCS, that represents the co-evolution of resources and processes. Mathematically, the basic idea is that a system consists of static parts — think, for example, of the hardware — and its dynamic parts — think of the software processes running on the hardware. This distinction is captured simply and directly by considering a process calculus that models the co-evolution of resources, the static parts, and processes, the dynamic parts:

$$R\,,\, E \xrightarrow{a} R'\,,\, E'$$

describes the co-evolution of resource $R$ and process $E$ when an action $a$ occurs. In particular, we assume that $R'$ is given as a function $f$ of $R$ and $a$, so that if $E$ is a process of the form $a : F$, which performs an $a$ and becomes $F$, then we have, as the basic step in the operational (or proof-theoretic) definition of the **SCRP**,

$$R\,,\, a : F \xrightarrow{a} f(a, R), F.$$

Other constructs, including parallel composition

$$\frac{R\,,\, E \xrightarrow{a} R'\,,\, E' \quad S\,,\, F \xrightarrow{a} S'\,,\, F'}{R \circ S\,,\, E \times F \overset{a\times b}{\to} R' \circ S'\,,\, E' \times F'}$$

and hiding, which forces a portion of the resources to be accessible only some chosen process,

$$\frac{R \circ S\,,\, E \xrightarrow{a} R' \circ S'\,,\, E'}{R\,,\, (\nu S)E \overset{(\nu S)a}{\to} R'\,,\, (\nu S')E'}$$

for a certain construcible action $(\nu S)a$ and some others, allow the (Turing complete) description of complex systems.

The basic theorem here is that there is a relation $\sim$ on resource-process pairs that is a bisimulation.

The many necessary technicalities required to set up and establish the properties of **SCRP** may be found in our writings.

The **SCRP** system is closely related to the DEMOS2k modelling tool (www.demos2k.com), itself a development of the DEMOS (Discrete Event Simulation On Simula ) system due to Birtwistle . Useful semantics of DEMOS can be given either in CCS or SCCS but the latter perhaps better captures DEMOS2k's treatment of resources. Thus **SCRP** may be seen as a natural mathematical generalization of the basis of DEMOS2k. DEMOS2k also embodies, via its treatment of resources, a natural representation of certain classes of queues. The representation of queue-theoretic ideas in **SCRP**remains to be explored. DEMOS2k is used both within HP Labs and in HP's businesses.

Along with SCCS comes a logic, usually called Hennessy-Milner logic, based on the following model-theoretic judgement:

$$E \models \phi,$$

read as 'process $E$ has property $\phi$', where $\phi$ is a proposition from the a logic containing, conjunction, disjunction, classical negation, and action-modalities $[a]$ — necessarily after $a$ — and $\langle a \rangle$ — possibly after $a$. The theory has been well-developed in a range of settings, for which provides a good starting point.

**SCRP** also comes with such a logic, derived from the resource semantics, with judgement

$$R\,,\, E \models \phi,$$

read as 'process $E$ has property $\phi$ relative to resources $R$'. In this logic, called **MBI**, $\phi$ may be built from the multiplicative and additive connectives of (Boolean) **BI**, together both additive and mutliplicative action-modalities, and additive and multiplicative first-order quantifiers.

The additive modalities are temporal, as in the logic for SCCS, but the multiplicative modalities are *spatial*, in the sense that their truth conditions depend on global modifications of the resources. The additive quantifiers have the usual classical first-order meaning, but the multiplicative quantifiers, like the modalities, refer to resource modifications. In the judgement $R\,,\, E \models \phi$, the multiplicative quantifiers capture the hiding of resources as described by the $(\nu S)E$ construct introduced above. The details are provided in [**?**].

The main theorem of note here establishes the relationship between operational equivalence and logical equivalence:

**Theorem 1** (bisimilar iff logically equivalent)**.** *For image-finite processes,*

$$R\,,\, E \sim R\,,\, F, \text{ for all } R, \text{ iff } E\,,\, R \models \phi, \text{ for all } R$$

## A.5 Proposed Programme of Research

### A.5.1 Introduction

The theoretical development of **SCRP** and **MBI** to-date, as described above, though well-motivated by applications, has been quite limited, with only the theorems above, together with a range of

exmaples, having been established. Before we can proceed to explore either detailed examples or prototype computer-based tools, or indeed the broader theoretical context of dynamical systems, a range of basic theoretical tools must be established.

### A.5.2 Some Theory

The theoretical development of **SCRP** and **MBI** to-date, as described above, though well-motivated by applications, has been quite limited, with only the theorems above, together with a range of exmaples, having been established. Before we can proceed to explore either detailed examples or computer-based tools, or indeed the broader theoretical context of dynamical systems, a range of basic theoretial tools must be established.

The first few things are relatively standard developments within process theory and process logic, and will provide and ideal starting point for the project.

**The (in)equational theory of SCRP processes**: it is evident that **SCRP** terms satisfy the usual basic algebraic laws, such as commutativity and associativity,

$$R\,,\, E \times F \sim R\,,\, F \times E$$

$$R\,,\, E \times (F \times G) \sim R\,,\, (F \times E) \times G,$$

and several others. But the substantive question concerns the *expansion theorem*. In SCCS, it is

$$\Sigma_{ij} a_i \times b_j : E_i \times F_j \sim (\Sigma_i a_i : E_i) \times (\Sigma_j b_j : F_j).$$

In **SCRP**, we do not expect to get a useful form for bisimulation. The main reason for this is that when we consider the constituent parts of a parallel composition we will have a particular allocation of resources to each of those parts. When we form the parallel composition we naturally form a (typically larger) compound resource, it is clear that this could have been divided in many ways other than that which we chose to do the original proofs of the behaviours of the sub-components. Whilst this observation does not matter when we are reasoning operationally and decomposing the structure, it is clearly important when we are reasoning equationally and forming terms by composition. The appropriate form in this instance is an *inequational* theory with the obvious extension of the standard expansion theorem for SCCS, with the caveat that, as a consequence of the potential ability to change the division of resource, the relationship will be one of *simulation*. This issue is closely related to **SCRP**'s representation of *asynchrony*.

For modelling purposes, there is by now a great deal of evidence that combining process-theoretic calculi with probabilities, priorities, and time leads to highly effective tools. For example, Tofts' WSCSS has found a wide variety of compelling modelling applications and, indeed, DEMOS2k is a probabilistic system. We propose to develop a probabilistic version of SCRP, to be called WSCRP, in a style similar to WSCCS. We will need to develop WSCRP's metatheory, addressing the questions of operational semantics, (in)equational theories, and bisimulation. Related work by Kwiatkowska and colleagues, on (asynchronous) calculi such as CSP , will be considered.

Turning to the logic, **MBI**, several developments are also required. First, to handle infinite state, we require least and greatest *fixed points*, that is formulæ of the form $\mu X.\phi X$ and $\nu X.\phi(X)$, for a propositional (*i.e.*, second order) variable $X$. This should be a straightforward adaptation of standard techniques to **MBI**, and will provide an ideal starting point for this part of the project. More substantially, in order to support *model checking* procedures and tools, we shall need to develop a *tableaux* system for **MBI**.

The theory of tableaux for basic **BI** is by now well-understood, with a susbstantial theoretical paper by Pym and colleagues at INRIA Lorraine, Nancy, and a prototype theorem prover, BILL, is available. This work, however, is all for **BI** with intuitionistic additives. Although we could formulate an intuitionistic version of **MBI**, its rôle as a system modelling and specifying logic is facilitated by having classical, or Boolean, additives. There are two difficulties to be overcome here.

First, the issue of suitable classes of models for Boolean **BI** for which completeness theorems are available. The class of models used for basic **BI**, based on preordered monoids (monoidal categories) is simply not general enought to handle classical, involutive negation. Solutions from relevance logic typically do not address the co-presence of both additive and multiplicative connectives, and linear logic lacks any convincing truth-functional semantics. Initial progress has been made by Yang, Galmiche, and others .

Second, the extension of such a semantics to **MBI**, with its action-modalities. Again, there is body of relevant work, such as , but our problem remains significant.

Finally, a tableaux system should provide a basis for understanding whether we can have a *sequent calculus* for **MBI**. Here the main issue will be the handling within a sequent calculus of the tableaux labelling constraints that enforce correct behaviour of the multiplicatives. Although this is straightforward for basic **BI**, the presence of the action-modalities is a significant complication. Work by Galmiche, Méry, and Pym , by Yang , and by Simpson will be helpful.

In the probabilistic setting, we will not seek to develop a 'probabilistic logic', such as Continuous Stochastic Logic corresponding to **MBI**. Rather, we will seek to develop WSCRP in such a way that we can retain our ability to capture properties of interest — such as separation properties of systems components — in **MBI**. For theoretical purposes, such as logical characterizations of bisimulation equivalence, constructs of the form found in PCTL may be of use.

### A.5.3 Examples, Tools, Evaluation

What are the natural 'resource-process architectures' for a range of examples? We will make a detailed exploration of the mathematical structure and properties of a range of examples drawn from topics such as commodity supercomputer architectures and large-scale IT systems. We will draw heavily upon HP Labs' ability to access the decriptive data for large-scale, implemented systems projects (see above).

DEMOS2k is a system simulation tool based on the Simula language and Birtwistle's modelling methodology. DEMOS2k can be explained semantically both in asynchronous and synchronous terms but, for our purposes, the treatment of resource embodied in the synchronous semantics is the more natural reference point. We will build (experimental) tools in the style (discrete event simulation) of DEMOS2k to implement, experimentally, the **SCRP** calculus and provide a model checking tool, for property verifcation purposes, based on the modal process logic **MBI**. We will seek to provide an explicit understanding of how such a tool can be used to implement conveniently systems models based on queue-theoretic and probabilistic models (WSCRP). That is, we will provide (experimental) tools to explore not only the components of the judgements $R, E \xrightarrow{a} R', E'$ and $R, E \models \phi$, and their integration, but also the stochastic aspects of their evolution.

In addition to DEMOS2k, other existing tools, such as the Concurrency Workbench will provide useful reference points and experience. Tofts, in particular, has significant experience with these tools and, moreover, has already supervised the developed at HP by an intern (Jonathan Hayman) of a model checking tool for a precursor to **SCRP-MBI**, developed with help from Pym with its theoretical aspects.

Several criteria will be important in evaluating our work: mathematical elegance, certainly, but also effectiveness in industrial-strength practice, and usability by time-pressed modellers.

### A.5.4 More Theory

Having established the discrete theory of resources and processes, as sketched above, we shall proceed to make an explicit connection with continous models.

The basic idea is that, corresponding to each atomic action $a$ in the calculus, there will be a function $[\![a]\!]$ in the space of resources. There are several steps required to make this precise. First, we will construct a *denotational semantics* for **SCRP**. The denotational meaning of a process $E$ can be defined by induction using 'synchronization trees', essentially as follows:

$$\begin{aligned}
[\![a : E]\!]_I R &\simeq \langle a, [\![E]\!]_I f(a, R)\rangle \\
[\![E \times F]\!]_I R &\simeq \int^{S,T} [\![E]\!]_I S \times [\![F]\!]_I T \times \mathcal{R}(S \circ T, R) \\
[\![E + F]\!]_I R &\simeq [\![E]\!]_I R \uplus [\![F]\!]_I R \\
[\![(\nu S)E]\!]_I R &\simeq [\![E]\!]_I (R \circ S)
\end{aligned}$$

for suitable choices of pairing, $\langle -, - \rangle$, and sum $\uplus$. Here $I$ fixes the basic signature of the model, and $\int$ is the co-end construction, used here to define Day's tensor product in 'presheaves' (over $\mathcal{R}$ as a monoidal category) valued in a suitable set-theoretic domain. We must also handle definitions in order to interpret recursive terms.

Subject to some refinements of this formulation, we will initially address two conjectures about this

semantics. The first should be quite straightforward:

**Conjecture 2.**

$$\text{For any } I,\ R,\ E \sim R,\ F \text{ implies } [\![E]\!]_I R = [\![F]\!]_I R$$

The converse, *i.e.*, *full abstraction*, is more challenging:

**Conjecture 3.** *There is a semantics* $[\![-]\!]_J$ *such that*

$$[\![E]\!]_J R = [\![F]\!]_J R \text{ implies } R,\ E \sim R,\ F$$

These two basic results will establish the basic correctness of the semantics. The first is essential, the second desirable.

### A.5.5 Dynamical Systems in $Set^{\mathcal{R}}$

Having established this functional, set-theoretic semantics, we can use it as a basis for investigating the continuous counterpart to our discrete modelling. The basic idea goes, in very simplifed terms, as sketched below.

We consider the case in which we have a language of actions $a$ for which modification is time-dependent; that is $[\![a]\!]_I = f(a, R, t)$. Suppose, for example, we have just two basic actions, $a$ and $b$, of interest, with modifications $A(R,t)$ and $B(R,t)$. $A$ and $B$ may be coupled as a differential equation, such as

$$\frac{\partial A}{\partial t} + \frac{\partial B}{\partial t} = C.$$

Observe that we have now moved to a time-dependent presheaf semantics of our calculus of resources and processes. Now we can investigate dynamical systems, described by equations such as the one above, in this space. Some questions for us to address arise immediately , such as:

- How are fixed points and limit cycles in the dynamical system related to fixed points in the logic?
- How can stability properties of the dynamical system be related to the the structural description of the systems and its logical properties/specification?
- How is the choice of topology in the semantical space reflected in the logical space?

### A.5.6 More Examples, Tools, Evaluation

We will, as before, study examples of systems for which both continuous and discrete models are valuable. Good choices might be load-surges in data networks (such as internet loading on 9/11), performance variation of large parallel systems because if 'compute' noise

Telephone (voice) networks (specifically, surges following faults); Data networks (specifically, surges in reaction to disasters); Evacuation of large buildings and/or urban ares under different types of threat; Performance variation on very large integrated circuits; Performance variation of large parallel systems because of 'compute' noise.

We will begin to consider the design of tools for modelling at the discrete-continuous interface. Such tools will allow choices in the design space of a discrete model of a system to be reflected as constraints in the continuous model, and vice versa. The evaluation criteria will similar to those for the discrete tools, but with much less exposure to practice.

We emphasize that any tools developed in this section — that is, for the discrete–continuous interface — will be for demonstration and proof-of-concept purposes only. A suitable programming environment will be provided by a computer algebra system, such as a Mathematica or Maple . In the event that we makew sufficent progress that we require access to source code, we will make use of the open source system Maxima . We may be able to draw inspiration from aspects of the PRISM system, developed by Kwiatkowska and others.

### A.5.7 Culmination and Speculation: Modelling Business Processes

Contracts between service IT service providers and their customers are regulated by Service Level Objectives (SLOs) and and Service Level Agreements (SLAs).

Taylor, Tofts, and Yearworth have introduced the notion of the Economic Value Principle (EVP) as a conceptual mechanism for coupling functional and transformational properties of compositional models of systems with their economic equivalents and so enabling the construction of contracts whose structures and properties closely reflect the system capabilites and service objectives.

Building on the mathematical technologies and tools to be developed in this project, our concluding exploration will be of the possiblity of developing systematic, logical and programming tools for designing SLAs for the deliverly of services by a system that conforms to a model openly agreed between the customer and the service provider.

Initial outputs might take the form of semi-automated procedures for constructing templates for SLAs from system models.